npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-025-01801-4

Generalizable image segmentation for microstructure characterization through integrated SEM and EBSD analysis

Check for updates

M. F. N. Taufique[1,4], Julia Nguyen[2,4], Julian D. Escobar[1,4], Mayur Pole[1], David Garcia[1], Tianhao Wang[1], Hrishikesh Das[1], Kathy Nwe[3], David Brown[3], Kenneth Ross[1], Donald R. Todd[2], Erin Barker[2], Eric Smith[2], Keerti S. Kappagantula[1] ✉ & Jenna A. Bilbrey[2] ✉

We demonstrate generalizable semantic segmentation using minimal ground truth data. Correlated scanning electron microscopy (SEM) images and electron backscatter diffraction (EBSD) measurements of friction-stir processed 316L stainless steel plates were used to train deep learning models for grain boundary segmentation. Secondary electron (SE) imaging taken at 10 keV correlated to EBSD-derived grain boundaries produced the best performing model. Notably, an ensemble of three models trained on a single SE image produced accurate segmentation over a series of backscatter electron (BSE) images of samples manufactured under different processing parameters, with a mean absolute error in grain size of 0.34 μm. The generalizability of the models likely results from the similar escape depths of the SE training input and the EBSD training output and the reduced probability of stored strain artifacts appearing in the image. This highlights the importance of considering the physical principles behind imaging to develop robust models for microstructure characterization.

Friction stir processing (FSP) is a form of solid-phase processing used to join and modify the microstructure of materials[1]. In this technique, a non-consumable rotating tool is plunged onto a metal surface, inducing friction and deformation heating, and subsequent plastic flow around the tool[2]. The FSP tool is then traversed along a desired length, producing a layer of refined grains underneath[3]. FSP has demonstrated the ability to join[4–6] two pieces together or modify[7,8] the microstructure. Ultrafine grain sizes can be achieved by controlling the heat input through fast processing speeds and slow tool rotation speeds[9]. The final stir zone microstructures benefit from grain boundary hardening (i.e., Hall-Petch hardening), which improves the strength of the processed material while minimizing degradation in the heat-affected zone[6,9]. FSP has also proven useful for repairing stress corrosion cracking and sensitization[10] and improving cavitation erosion resistance[7] of 304/304 L austenitic stainless steels[10].

The microstructure of a material plays a crucial role in determining its physical, mechanical, and functional properties. The relationship between mechanical properties, such as hardness and grain size is well-established, with the Hall-Petch relation demonstrating an inverse dependence of material strength on grain size[11]. In FSP, grain size is largely influenced by the processing parameters employed during manufacture. Garcia et al. established a relationship among processing temperature, grain size, and hardness for 304 stainless steels, where lower processing temperatures produced smaller grain sizes, leading to higher hardness values[12]. Notably, localized variations in microstructure can occur for a static set of processing parameters. For instance, Wang et al. observed variations in grain size in the center versus the bottom of the stir zone, which correlated with variations in hardness[13]. The cause of such variation is thought to be due to the presence of temperature and strain gradients through the material as a result of the tooling geometry. Liu et al. demonstrated the complexity of the microstructural evolution at different processing positions[14]. The unprocessed material in front of the tool undergoes rapid deformation, leading to grain fragmentation via the generation of low-angle grain boundaries (LAGBs) and the deterioration of pre-existing annealing twin boundaries. However, within the stir zone microstructure, the combination of strain, strain rate, and temperature leads to dynamic recrystallization, reducing LAGBs and promoting twinning[5,14–16].

The final FSP microstructure of austenitic stainless steels is complex, and therefore, electron backscatter diffraction (EBSD) is needed to achieve a complete understanding of the crystallographic evolution induced by the process. For example, the coexistence of LAGBs and high-angle grain

[1]Energy & Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, USA. [2]National Security Directorate, Pacific Northwest National Laboratory, Richland, WA, USA. [3]Computing & Information Technology Directorate, Pacific Northwest National Laboratory, Richland, WA, USA. [4]These authors contributed equally: M. F. N. Taufique, Julia Nguyen, Julian D. Escobar. ✉e-mail: ksk@pnnl.gov; jenna.pope@pnnl.gov

boundaries (HAGBs), coincidence site lattice (CSL) Σ3 annealing twin boundaries, and dense dislocation walls results in a rich variety of observable features when using conventional electron imaging modes, such as secondary electron (SE) and backscatter electron (BSE) microscopy. The inherent microstructural variability within FSP microstructures requires large area EBSD characterization on the scale of millimeters[4,5,14,17]. However, accurate EBSD data is closely related to a fine beam step size at the nanometer scale employed during data collection, which in turn requires long collection times. This is especially true when large areas need to be characterized to understand the evolution of microstructure in processed specimens. In the case of solid-phase processed austenitic stainless steels, a high-quality EBSD map takes longer to capture than a comparable SE/BSE image over the same area. The quality of EBSD data is strongly influenced by surface quality after standard metallographic preparation, which is a labor-intensive and expertise-dependent task. Efforts in high-throughput experimentation and quality assurance/quality control (QA/QC) have motivated exploration into advanced image analysis techniques, with a particular focus on deep learning, to streamline the process of microstructure characterization.

Deep learning offers advanced toolsets to enhance and automate the analysis of microscopy images[18–20]. Semantic segmentation, in particular, has begun to show use for high-throughput materials characterization[21]. Roberts et al. introduced a semantic segmentation model, called DefectSegNet[22], to identify dislocation lines, precipitates, and voids in transmission electron microscope (TEM) images of structural alloys. Patrick et al. applied U-Net to detect grain boundaries in bright-field TEM images containing high intragranular contrast[23]. Shen et al. used phase maps from EBSD to train a U-Net model to segment ferrite, martensite, and retained austenite regions in scanning electron microscope (SEM) images of dual-phase steel[24]. Notably, their model was robust against varying imaging modalities, qualities, and magnifications. Hirabayashi et al. trained segmentation models on 3D-SEM SE images that were able to identify boundary regions along the depth direction[25]. However, the inability to directly correlate the escape depth between the SEM and EBSD data collected from a sample necessitates manual labeling of 3D-SEM training sets to identify the various features of the microstructures.

Though standard segmentation model architectures can be applied to microscopy data, the collection of large volumes of labeled microscopy images on which to train a model is often an issue. To address this, Stuckner et al. compiled a large dataset, called MicroNet[26], containing over 100,000 labeled images obtained from TEM, SEM, and optical microscopy. They demonstrated that the finetuning performance of many standard segmentation architectures improved when pretrained on MicroNet; in some cases, only a single image was necessary to finetune a model.

In our previous work, multiple segmentation architectures were employed on SEM images of 347H stainless steels manufactured by cold rolling the plates, followed by annealing. Training labels were provided by EBSD, which were used to segment grain boundaries and quantify grain size distributions in the correlated SEM-EBSD images[27]. Annealing of the samples produced a clear visual distinction between the grains and grain boundaries, leading to impressive model performance. However, FSP produces a more complicated microstructure, where the austenitic grains are substantially finer and contain considerable amounts of LAGBs and dense dislocation walls.

Here, we expand upon our prior segmentation work with a focus on characterizing highly strained FSP microstructures. FSP microstructures present unique challenges to quantitative analysis. Internal strain in FSP microstructures results in high intragranular contrast in SEM-BSE images, which complicates both traditional analysis and the hand-labeling of grain boundaries. To address impediments to strained image analysis, we examined the coupling of different modalities (BSE and SE) to EBSD measurements to establish high-quality labeled data on which to train semantic segmentation models to identify grain boundaries. We then used the ensemble of models trained on the best-performing modality to segment a series of BSE images of samples manufactured with different FSP parameters. We found that despite the modality of the training data being different from that of the new test data, the model provided accurate predictions. Our results showed that the physics-based processes of microscopy imaging were key to determining the 'goodness' of the training data, which was crucial for model performance.

Results

EBSD and SE/BSE overlays

Image segmentation can be used to accelerate grain size analysis over large areas based on SE/BSE imaging. It is well-known that EBSD measurements are more time-consuming than SE/BSE and that the data acquisition time is closely related to the selected step size. For example, in our training data, a map collected in 50 nm steps covering an area of approximately 450 μm^2 required approximately 17 min of EBSD data collection. A similar area can be imaged in a few seconds via SE/BSE, but the quality varies depending on the image resolution and pixel dwell time (i.e., time consumption). The SE/BSE images in this work have a native resolution of 2560 × 2048 pixels, covering an approximate area of 490 μm^2 (which allowed for cropping and alignment relative to the EBSD images), and were collected over 160 s (2 min and 40 s), resulting in a potential 14–15 min time savings per image. This work aims to understand the effect of SE and BSE image modes, as well as the effect of different acceleration voltages, on image segmentation and model accuracy. An additional motivation for this work is the relative widespread accessibility of SE/BSE imaging over EBSD, as the former is available in virtually all SEM instruments at academic and research institutions.

The first step in developing a segmentation model for grain boundary detection is to obtain high-quality, properly labeled microscopy data on which training can be conducted in a controlled and reliable fashion. However, there are fundamental aspects to the generation of SE and BSE images that first need to be clarified. Although these concepts are well-known within the microscopy community, materials scientists and data scientists who are less familiar with microscopy may find these clarifications useful for future engagement with AI/ML for microscopy. First, we briefly examine the physical mechanisms behind image collection modes and acceleration voltages. Readers are encouraged to review other sources for a more detailed understanding of electron-matter interactions in SEM[28]. A quick summary of image generation modes is provided as follows:

- *BSE:* Backscatter electron images are produced by detecting elastically scattered electrons reflected as the beam interacts with the sample. The contrast within a BSE image is strongly sensitive to composition (atomic number), crystal orientation, and crystal defects of the sample material due to diffraction and channeling[29].
- *SE:* Secondary electron images are produced by detecting low-energy electrons resulting from inelastic scattering between the electron beam and the sample. Secondary electrons originate at or near the surface of the sample and, therefore, are mainly sensitive to topography and less sensitive to atomic number variations in the sample or the crystallographic orientation of grains. However, crystallographic contrast can still be observed in SE imaging because backscattered electrons exiting the sample can induce other secondary electrons near the surface[30].
- *EBSD:* Electron backscatter diffraction images are generated via diffraction of the backscattered electrons. Different from BSE imaging, the EBSD sample is tilted 70° relative to the horizon, so the BSE yield signal changes from isotropic (no tilt) to strongly forward peaked (tilt), increasing the interaction path and the outcoming signal[31]. As the backscattered electrons spread underneath the surface and interact with the sample, constructive diffraction occurs with the crystal planes of the sample that satisfy Bragg's law. The diffraction patterns, known as Kikuchi lines, are recorded by the EBSD detector and compared against a database, allowing for microstructural and crystallographic identification.

In the current study of FSP 316L stainless steel samples, image contrast variations in both SE and BSE modes due to atomic number can be disregarded since austenite is a solid solution that reflects the average

composition of the steel. Contrast variations due to orientation and electron channeling become of prime importance, especially in fine-grained face centered cubic (FCC) solid solutions produced via solid-phase processing during FSP. Figure 1 summarizes the variability of the microstructural features of FSP 316L austenitic stainless steel observed via BSE as a function of the acceleration voltage in steps of 2 keV. In addition, BSE and SE modes are compared at acceleration voltages of 10 and 20 keV, as well as against EBSD data obtained at 20 keV.

Increasing the acceleration voltage increases the contrast (i.e., signal-to-noise ratio) between microstructural features in the BSE images, especially at and above 6 keV. Interestingly, certain microstructural features appear or disappear as the acceleration voltage is increased. This effect is illustrated by the red rectangles in Fig. 1, which highlight an austenitic grain containing annealing twins that are only observable below 12 keV. Furthermore, the intragranular features within white squares, which are not associated with HAGBs or LAGBs, seem to fluctuate as the acceleration voltage is changed. These regions are associated with dense dislocation walls separating small domains inside grains, with misorientation angles smaller than 2°. SE images taken at 10 and 20 keV are comparatively less noisy and have milder contrast compared to their BSE counterparts. Nonetheless, these images are still sensitive to crystallographic contrast, effectively revealing grain and twin boundaries, while being less sensitive to dense dislocation walls. Therefore, SE images are an alternative imaging option that contains the features of interest in this work while limiting the contrast of small intragranular misorientation.

Next, our analysis of EBSD step size and imaging accelerating voltage raises an important aspect regarding the inherent variability of electron microscopy images as a function of imaging conditions. The variability implies that there is not a single EBSD, SE, or BSE image that defines the ultimate 'ground truth' microstructure of a sample. However, a compendium of multiple images represents the same microstructure seen by EBSD, which alters our perception of the ground truth of a sample's microstructure. Consequently, if a ground truth image from microscopy is required for training deep learning models, it must be accompanied by an adequate label or metadata, describing the measurement conditions that were used to generate such image.

First, to define a quality crystallographic ground truth and to understand the time consumption associated with each measurement, we

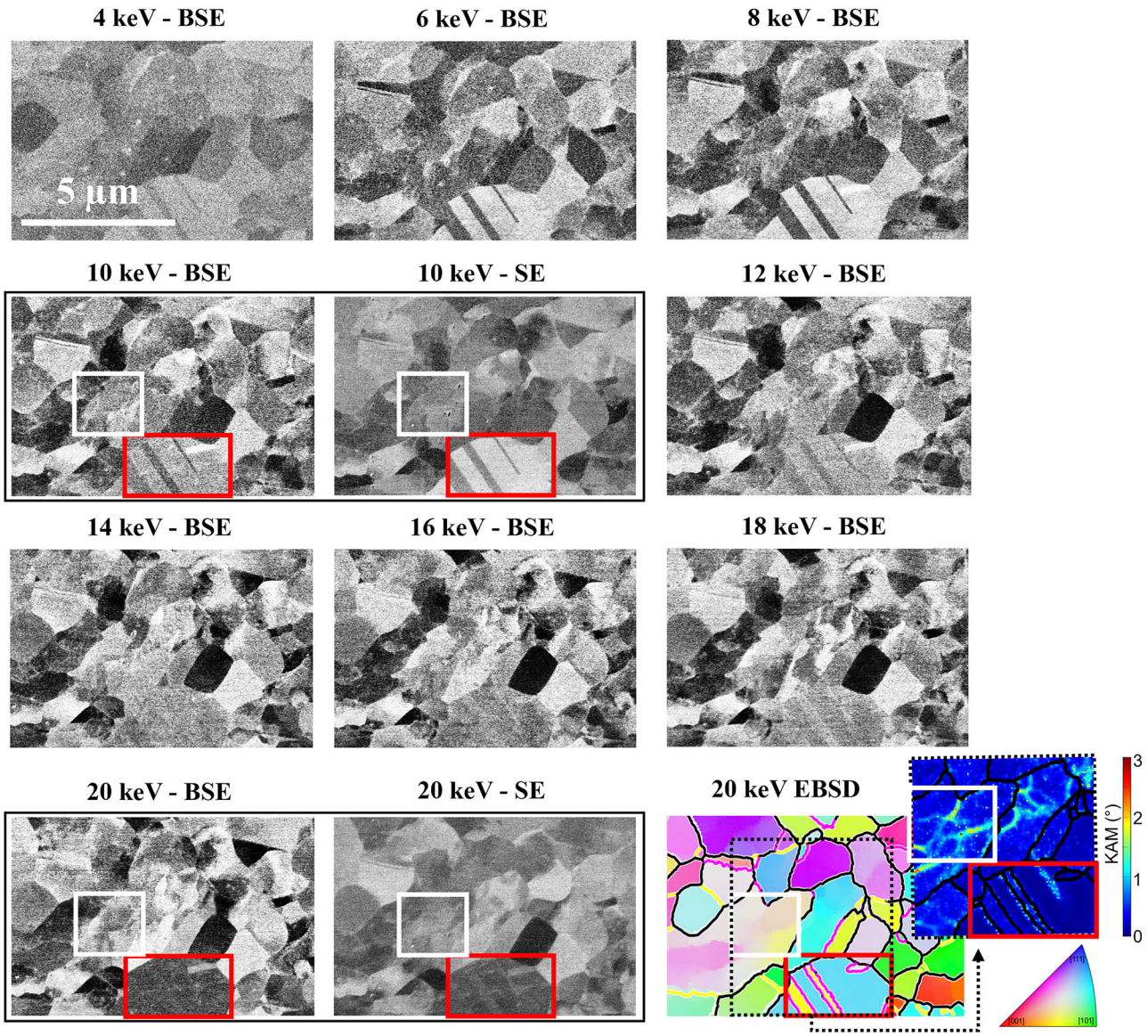

Fig. 1 | Micrographs collected using SE and BSE at different acceleration voltages from 4 to 20 keV, compared to an EBSD crystallographic map collected at 20 keV. EBSD analysis shows the inverse pole figure (IPF) map in the Z direction overlaid with grain boundaries colored according to grain and subgrain crystal orientations (HAGBs: black; LAGBs: yellow; annealing twins: magenta) and the kernel average misorientation (KAM) map, which details the presence of dense dislocation walls inside some of the grains. All boundaries in the KAM map are shown in black to better highlight the positions of the dense dislocation walls.

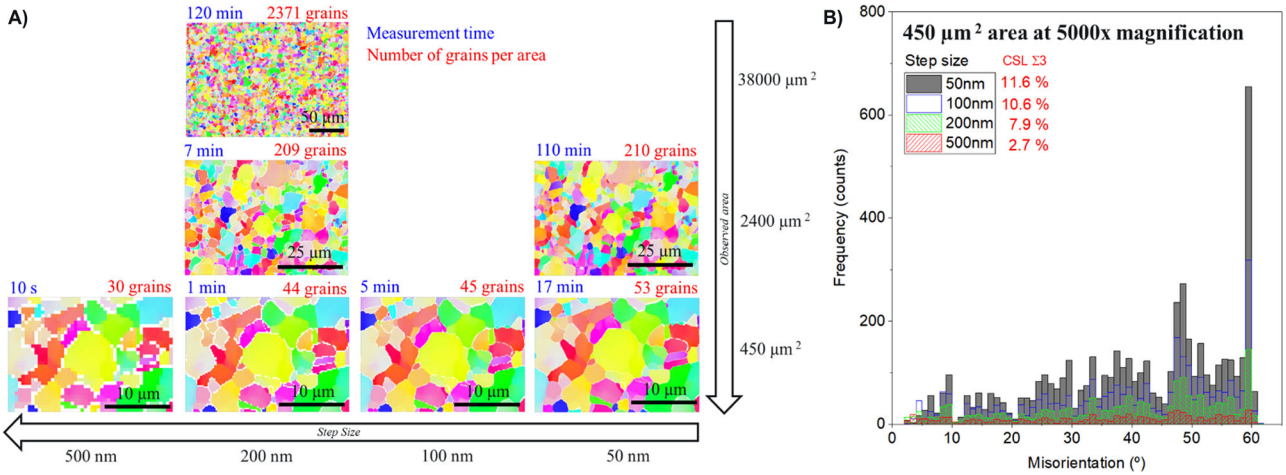

Fig. 2 | Parametric exploration of EBSD collection. A Relationship between measurement time, observation area, and step size and **B** frequency of misorientation angles over a fixed 450 µm^2 observation area at ×5000 magnification as a function of step size.

explored different step sizes and areas of interest, as summarized in Fig. 2. Our main objective was to obtain EBSD maps that contain sharp grain boundaries, in a reasonable measurement time, and with minimum loss of information. As seen in Fig. 2A, time consumption increases both as the step size is reduced (higher pixel densities) and as the area of interest is increased. The effect of step size on the number of measured grains and on the accuracy of boundary identification, particularly for CSL Σ3 twin boundaries, is shown in Fig. 2B. Coarse step size yields fast results but at the cost of reduced twin boundary detection and grain count. More details on the quality of the reconstructed grain boundaries can be seen in the supplementary information (Fig. S1). For model training purposes, we selected a step size of 50 nm to maximize the quality of our crystallographic ground truth, requiring an elapsed time of 17 min per ~450 µm^2 area of interest. Grain boundaries were carefully reconstructed following the protocols described in the Methods section, aiming to obtain continuous grain boundary skeletons that fully enveloped every identified grain. Examples of discontinuous grain boundaries and successful post-processing are shown in the supplementary information (Fig. S2).

Acknowledging the inherent variability of SE and BSE images, we opted to define our ground truth as the crystallography-based data generated via EBSD at a fixed 20 keV acceleration voltage and 50 nm step size. During processing of the EBSD data, we reconstructed the grain boundaries and deconvoluted this information into skeleton-like grain boundary maps.

Semantic segmentation requires labeled training data, meaning that the input image must have a corresponding segmentation map with a pixel-to-pixel match. To create labeled training data for microscopy images, previous studies have used hand-drawn segmentation maps[22,23,32]. FSP causes the formation of numerous LAGBs and dense dislocation walls in the 316L austenitic stainless steel microstructure. Consequently, BSE images are of high contrast (orientation and electron channeling effects), compromising accurate manual identification of grain boundaries. Therefore, we performed sequential, correlated SEM (SE/BSE) and EBSD measurements to produce labeled training data.

To create the training data, the ground truth grain boundary map (model output) was obtained from the EBSD boundary reconstruction. LAGBs, HAGBs, and twins were grouped into a single grain boundary class. Sample images are shown in Fig. 3. The registration of the SEM images and grain boundary maps is complicated due to stage tilting and trapezoidal distortion[33] in the EBSD image relative to SE/BSE, requiring specialized post-processing procedures. Differences in tilt angles between SEM and EBSD lead to foreshortening of grains and varying interaction volumes, while differences in magnification and working distance lead to varying image resolution, and differences in accelerating voltage and beam current lead to varying probe sizes—all of which require correction to obtain an accurate pixel-to-pixel match[34]. We first registered the fiducial marks in the EBSD grain boundary map and SEM images, then manually adjusted the EBSD grain boundary map to obtain pixel-to-pixel correspondence. The addition of markers greatly improved the success of obtaining a pixel-to-pixel correspondence by maximizing the spatial overlap between observation areas (see supplementary information Fig. S3). This process resulted in four sets of training pairs (see the Methods section for more details).

Our SEM-EBSD registration approach is similar to that of Shen et al.[24], who also used correlated SEM-EBSD measurements combined with manual adjustment of the EBSD map. Notably, while their segmentation models trained on this approach were able to accurately distinguish austenite and martensite phases in dual-phase steel, the models were not able to determine the exact locations of grain boundaries. This limitation was presumed to arise from the fine, indistinct, and "somewhat fuzzy" boundaries between phases.

Training and evaluation of segmentation models

Table 1 shows the average performance of the three individual UNet++ models for each modality and model pretraining/loss scheme. Specifically, the image modality label presents the information on the type of image (BSE or SE) and the accelerating voltage at which the image was collected that was used for training the UNet++ models. The F_1 scores, HD95, and mean absolute error (MAE) in the mean equivalent circle diameter (ECD) for individual models trained on each modality are presented in supplementary information Tables S2–S4.

Comparing across image modalities, we see that models trained on the SE image taken at an accelerating voltage of 10 keV (SE 10) performed best across all models over all three metrics. Across each modality, models pretrained on MicroNet outperformed those pretrained on ImageNet. The benefit of the addition of topoloss to the loss function (denoted TopoDICE) is unclear, with performance improving and worsening across different metrics and different training images. It appears that TopoDICE enhances accuracy in ECD when model performance is better, but decreases accuracy if a certain performance threshold cannot be met with DICE alone. The topological loss rewards conformity in the number of continuous, enclosed areas in the ground truth and predicted grain boundary maps, without considering the actual location of the grain boundary pixels. Conversely, DICE rewards pixel-level overlap and does not consider continuity. Therefore, we hypothesize that if pixel-level overlap cannot be accurately learned, rewarding continuity only further decreases accuracy.

For the best-performing modality (SE 10), TopoDICE reduced the MAE in ECD from 0.68 to 0.57 µm but gave the same average F_1 score of 0.62, though the average HD95 slightly increased from 26.3 to 28.7 pixels. Because our target task was to characterize grain structure, we weighted the

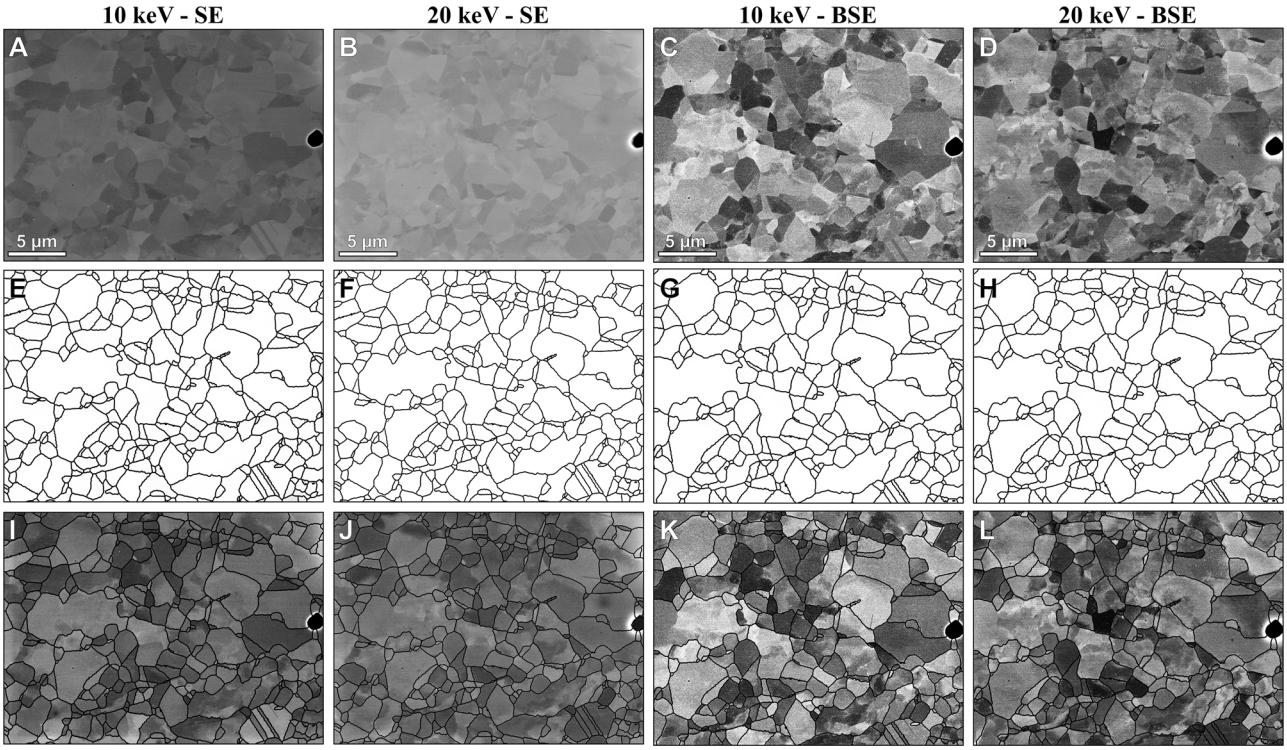

Fig. 3 | SE and BSE images, their corresponding EBSD-generated grain boundary maps and the overlay of the grain boundary maps on the SEM images at various accelerating voltages. A–D SEM images, **E–H** EBSD-generated grain boundary maps, and **I–L** overlays after registration.

Table 1 | Average three-fold cross-validation performance of models trained on different image modalities and with different pretraining/loss schemes

	F_1 Score			HD95 (pixels)			MAE in ECD (µm)		
Image modality	ImageNet DICE	MicroNet DICE	MicroNet TopoDICE	ImageNet DICE	MicroNet DICE	MicroNet TopoDICE	ImageNet DICE	MicroNet DICE	MicroNet TopoDICE
BSE 10	0.56	0.58	0.57	49.4	44.5	48.3	1.49	1.42	1.75
BSE 20	0.50	0.51	0.50	55.7	55.2	53.7	1.32	1.10	1.23
SE 10	**0.60**	**0.62**	**0.62**	**30.6**	**26.3**	**28.7**	**1.08**	**0.68**	**0.57**
SE 20	0.52	0.52	0.53	35.9	36.1	34.9	1.86	1.15	1.29

Bolded entries denote the best value for each metric within the set of image modalities. Underlined entries denote the best value for each metric within the set of pretraining/loss schemes

MAE in ECD higher than HD95 and, therefore, consider our best set of models to be the MicroNet/TopoDICE models trained on the SE 10 keV image.

To understand why model performance was highest when trained on the SE images obtained at an acceleration voltage of 10 keV, we reviewed the concept of interaction volume of the sample during imaging and image contrast. First, in the case of backscattered electrons, Monte Carlo simulations (provided in supplementary information Fig. S6) were performed for an equivalent 316L stainless steel solid solution and a beam normal to the surface. The estimated maximum penetration depth of backscattered electrons was approximately 17 ± 4 nm, 58 ± 10 nm, 121 ± 18 nm, and 190 ± 29 nm for beam energies of 5 keV, 10 keV, 15 keV, and 20 keV, respectively. During EBSD, however, the sample was tilted to 70° relative to the horizon, which reduces the interaction volume to 50–100 nm[35]. Based on this, an electron image that pairs with the EBSD map should be collected at a reduced acceleration voltage to reach a similar interaction volume. This is readily evident for BSE images but should also be considered for SE images, which are still mildly sensitive to crystallographic contrast. This is consistent with our observations in Fig. 1, where the best visual match between BSE/SE and EBSD information was observed for acceleration voltages below 12 keV, i.e., nearly half of the EBSD acceleration voltage of 20 keV.

Initially, the SE-EBSD pair may seem counterintuitive because of their differing scattering mechanisms. However, the reduced sensitivity of SE images to crystallographic contrast and electron channeling contrast, along with the shallower interaction volume relative to BSE, results in a more suitable image pair. One persisting limitation, even for SE-EBSD pairs, is related to the contrast variations caused by geometrically necessary dislocations. Although dense dislocation walls were excluded during the grain boundary reconstruction protocol, such regions are still present in the EBSD data and can be better highlighted via kernel average misorientation (KAM) analysis. Comparatively, regions of the microstructure inside the white rectangles in Fig. 1 show contrast variations associated mainly with misorientation build-ups around 2°. These regions can still be a source of false positive identifications by the segmentation models, especially if the SE image is acquired at a high contrast condition, and can lead to artificially fine grain size predictions.

Evaluation of out-of-distribution samples

Because our models were trained on images from a sample produced using a single set of FSP conditions, we were interested in investigating the ability of the models to accurately segment images from samples produced under different FSP conditions. Segmentation models for microscopy images are

known to generalize poorly to out-of-distribution (OOD) images due to differences in a variety of imaging and material parameters[36]. However, it is crucial for a segmentation model to be able to perform accurately with OOD images not used during training to improve its applicability across samples produced at different processing conditions. To assess the OOD performance of our models, we examined their performance in segmenting a set of 20 BSE images with differences from the training set in terms of both material processing conditions and imaging parameters. Specifically, the OOD images segmented by our models were manufactured using FSP conditions different from those used to process the sample from which the training image data was obtained. Additionally, the OOD images were also collected at different microscopy parameters, namely a different instrument, a different instrument operator, and a different modality. Table 2 gives the processing conditions, and Table 3 gives the imaging conditions. The only commonality among the OOD images and all training images is the material (316L stainless steel). The BSE training sets share the same imaging modality with the OOD set, and the BSE 20 keV training set also shares a common accelerating voltage.

Table 2 | FSP processing conditions for the training and out-of-distribution (OOD) samples used in this study

Sample set	Temperature (°C)	Traverse velocity (mm/min)	Force (kN)
Training	850	25.4	46.7
OOD	800	101.6	47.0
OOD	800	50.8	47.0
OOD	900	76.2	47.0
OOD	900	50.8	47.0

Table 3 | Comparison of imaging parameters across training and OOD images

Image set	Imaging modality	Accelerating voltage (keV)	Resolution (μm/px)	Observation area (μm^2)
Training	EBSD	20	0.016	450
Training	BSE	10	0.010	490
Training	BSE	20	0.010	490
Training	SE	10	0.015	770
Training	SE	20	0.015	770
OOD	BSE	20	0.027	1200

Figure 4 shows a BSE image from the OOD set overlaid with segmentation maps for the MicroNet/TopoDICE model trained on the SE 10 keV image, along with the corresponding grain detections. This image demonstrates the poor grain boundary closure for individual sets, and the improvement gained with ensembling. The individual models tended to produce segmentation maps with gaps in grain boundaries, which leads to erroneous grain detection and artificially increases the measured grain diameters. Ensembling the predictions by summing segmentation maps from the three models trained on the same modality led to improved grain boundary closure and, thus, grain detection. Supplementary information Tables S5 and S6 give the mean number of grains and error in ECD for each modality training set and ensemble. In each case, ensembling recovers more grains, which improves the accuracy of grain diameter measurements. Going forward, we applied ensembling for each model training modality when predicting segmentation maps for the OOD images.

We did not have pixel-to-pixel alignment between BSE images and EBSD measurements of the OOD samples. Thus, we validated model performance through comparison of the mean ECD determined from the ensembled segmentation maps and EBSD measurements. It should be noted that a perfect match in ECD between the two modalities is not expected due to differences in interaction volume, as discussed previously, as well as measurement technique. For instance, BSE images show contrast between grains, subgrains, and regions surrounded by dense dislocation walls, but these cannot be accurately categorized individually. Conversely, EBSD can provide such differentiation based on misorientation analysis.

Humphries et al. discussed the reasons behind the mismatch between grain size measurements between light optical microscopy, SEM imaging, and EBSD in weakly and strongly texturized aluminum[37]. Strongly

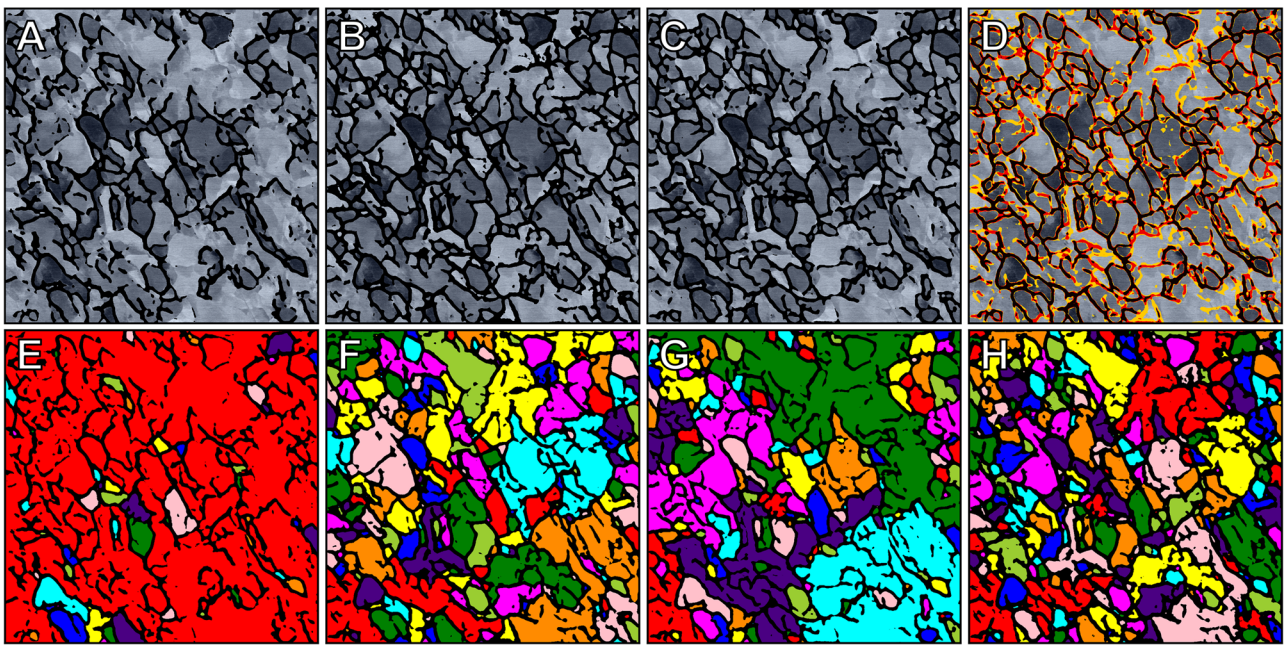

Fig. 4 | Comparing single model predictions to the ensembled prediction. **A**–**D** BSE images and overlaid segmentation maps produced by the UNet++/MicroNet/TopoDICE/SE 10 keV model for an OOD sample and **E**–**H** corresponding grain detections for **A**, **E** training set 1, **B**, **F** training set 2, **C**, **G** training set 3, and **D**, **H** the ensembled output of all training sets.

texturized microstructures containing high densities of LAGBs tend to show a smaller grain size based on imaging techniques, especially when images are sensitive to crystallographic contrast[37]. This occurs as all measurable boundaries contribute to grain size calculations via the line intercept methodology. In more randomized and recrystallized microstructures with low densities of LAGBs, optical, SEM, and EBSD based calculations tend to agree.

Table 4 gives the overall MAE in ECDs obtained from the ensembled segmentation maps over the full OOD sample set. The MicroNet/TopoDICE model trained on the SE 10 keV image gives the lowest MAE of 0.34 μm, followed by the MicroNet/DICE model trained on SE 10 keV of 0.40 μm. The MAEs for the MicroNet models trained on BSE 10 keV and 20 keV images were extremely high due to the drastic underprediction of grain boundary pixels.

Based on the MAE, it appears the ImageNet/DICE model provided more consistent, albeit less accurate, predictions across training image modalities. However, further examination revealed that the ImageNet/DICE model produced a narrow range of ECD values across the OOD samples. Figure 5A shows the individual predictions on OOD samples across models trained on the SE 10 keV images, and Fig. 5B compares ECD distributions for the models and EBSD. ECD distributions for the remaining training set modalities are given in supplementary information Table S7. EBSD-determined ECDs range from 1.47 to 4.68 μm, while ImageNet/DICE/SE 10 keV ECDs range from 3.07–4.40 μm, which are within the range of the grain sizes obtained from the EBSD 'ground truth'. The MicroNet models more closely reproduce the expected ECD distribution, especially at lower grain sizes.

Despite training on the SE 10 keV image, the models successfully transferred their learning of SE images to BSE images taken using different microscopes, different imaging settings, and by different operators of samples processed under different conditions with a wider range of mean grain sizes. From this observation, we can conclude that carefully considering the physical properties underlying the collection of the training data allows the generation of segmentation models that can accurately analyze images collected from different samples using different imaging modalities.

Table 4 | MAE in the ECD (μm) from the ensembled segmentation maps predicted for OOD image set against that measured by EBSD in the corresponding region

	Training set modality			
Model	**BSE 10 keV**	**BSE 20 keV**	**SE 10 keV**	**SE 20 keV**
ImageNet/DICE	2.04	2.49	0.87	0.85
MicroNet/DICE	19.40	11.64	0.40	2.07
MicroNet/TopoDICE	21.00	9.59	**0.34**	4.37

The bolded and underlined entry indicates model with the lowest MAE.

Training involves learning of pixel-to-pixel correlations between the input (SE) and output (EBSD) data, while prediction is validated by mean grain size rather than exact pixel overlap. BSE imaging indeed captures grain boundaries, though at a deeper interaction volume than SE or EBSD. We expect grains at the same location to have similar ECDs across the approximately 250 nm depth captured by BSE compared to the approximately 20 nm depth captured by EBSD.

Discussion

In this work, we developed an approach to segment microstructural images obtained from FSP 316L stainless steel samples that supports predictions across multiple imaging modalities and various manufacturing parameters. We trained UNet++ models with a single correlated SEM-EBSD image pair to predict grain boundaries and grains in SEM images, which were then used to calculate the ECD. Notably, an ensemble of three models trained on a single SE (10 keV) image performed well over a series of BSE images of samples manufactured with different FSP parameters, giving an MAE in grain size of 0.34 μm for samples with ECDs of 1–4 μm, as determined from EBSD.

The striking ability of the model trained on SE images to accurately segment and provide ECD statistics from BSE images likely results from several factors, such as (1) the interaction volume of the training input (SE 10 keV) closely matching that of the training output (EBSD), (2) the reduced level of noise in the SE image compared with BSE, and (3) the limited contribution of small intragranular misorientation features such as dense dislocation walls. Although BSE and EBSD may seem like the logical pairing for this type of training effort (as both are based on backscatter electrons), it became apparent that SE imaging provided the optimal tradeoff between mild grain/subgrain boundary contrast with limited susceptibility to stored strain.

This study presents a framework for developing segmentation models for images of complex microstructures seen during solid-phase processing of metallic specimens. The key finding from this study is that the purposeful collection of training data results in segmentation models can be effectively applied to materials processed by the same approach but under different conditions despite changing the imaging modality used for prediction.

Model transfer to other dynamically recrystallized FCC systems containing annealing twin boundaries (copper, nickel, super nickel alloys, etc.) manufactured by other solid-phase processing methodologies (hot forming, forging, friction extrusion, etc.) can be reasonably expected. However, additional work is needed to examine transferability to more complex microstructures, such as for aluminum or magnesium, or for manufacturing methods that lead to columnar grains, such as casting, welding, and additive manufacturing.

Methods

FSP samples

FSP was performed at Pacific Northwest National Laboratory on a custom Manufacturing Technology, Inc. friction stir welding machine. The specimens for this study are 316L austenitic stainless steel plates

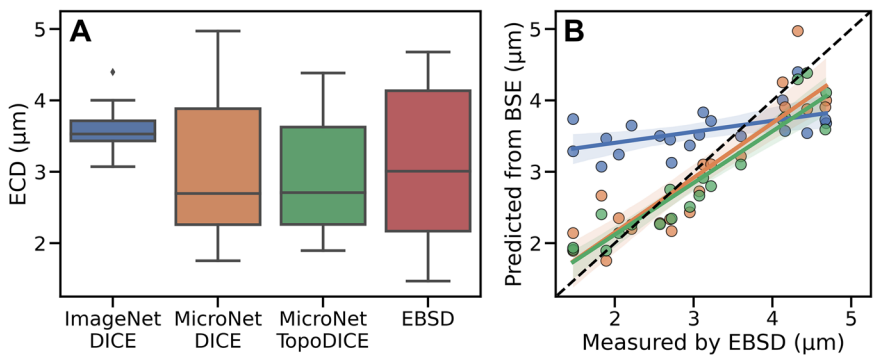

Fig. 5 | Comparing ECD predictions made by various models and by EBSD. A Distributions and **B** individual ECD predictions on the set of OOD BSE 20 keV images using the SE 10 keV model ensembles, compared against that determined by EBSD. The colors in (**B**) match those in (**A**), and the dashed black line indicates the ideal fit.

(152 × 330 × 10.3 mm) treated by FSP. All samples were processed under different tooling conditions, given in Table 2. A full description of the FSP setup is given in ref. 12. A single sample was used for training and in-distribution testing. Four samples were used strictly for testing the trained models on samples that are considered out-of-distribution (OOD) of the training data due to the different processing conditions used to manufacture the samples and the different microscope parameters used to collect the images, given in Table 3[12].

Training data

After FSP, samples were obtained from the processed 316L plates and polished for microstructural analysis. All samples were prepared using standard metallographic polishing, ending with 0.02 μm colloidal silica for approximately 6 h. No chemical etching was used in this work. EBSD data were collected using a JEOL 7600 SEM, equipped with an Oxford Synergy CMOS EBSD detector. An acceleration voltage of 20 keV, a step size of 50 nm, and a tilt angle of 70° were used in all measurements. Platinum fiducial markers were deposited on the stir zone of the four OOD samples at five different locations per sample. A pixel-to-pixel correlation between SE, BSE, and EBSD was possible by maintaining a constant region of interest using fiducial marks visible in all imaging modes to have a spatial resolution and overlap on SEM and EBSD.

BSE images were collected with a JEOL IT500 HRLV field emission scanning electron microscope. The region of interest was fixed, and images were collected in BSE mode between 4 keV and 20 keV with steps of 2 keV at 0° tilt, using a fixed resolution of 3600 × 3000 pixels, a standard probe current at a level of 75%, a working distance of 10 mm, and a magnification of ×5000. All SE images were collected with a FEI Quanta 3D FIB Scanning Electron Microscope using a resolution of 2048 × 1886 pixels, at 10 or 20 keV, 0° tilt, a standard probe current of 22.6 nA, working distance of 10 mm, and magnification of ×10,000. Note that the observation area from EBSD training data was set as the region of interest over which BSE/SE were collected. BSE images were collected over a slightly larger area to allow for alignment and cropping. SE images covered a larger area relative to BSE due to a difference in image form factor and to allow for cropping and alignment of the same region of interest.

For EBSD, data analysis was performed using MTEX 5.8.2 toolbox[38] in Matlab R2020b. A five-pixel neighbor clean was used to denoise random electronic noise and local zero-solution zones. Then, a misorientation threshold of ω = 15° was used to identify HAGBs, following the conventions used by Humphrey and colleagues[37,39]. LAGBs were identified for a misorientation window of 2° < ω ≤ 15°. The remaining intragranular boundary information was classified as dense dislocation walls, i.e., ω ≤ 2°, and was excluded from the LAGB and HAGB reconstruction sequence. Finally, annealing twin boundaries, i.e., CSL Σ3, were reconstructed using the methodology proposed by Patala et al.[40]. These were reconstructed by identifying boundaries at 60° misorientation oriented about <111>. After boundary reconstruction and identification, LAGBs, HAGBs, and twins were deconvolved, labeled, and then extracted to generate independent boundary-specific skeleton-like plots (see supplementary information Fig. S4).

Out of distribution (OOD) data

Prior to microstructural characterization, platinum fiducials were deposited on the stir zone of the four OOD samples, shown in supplementary information Fig. S5, to ensure capture of the stir zone at various locations during imaging. BSE images were captured at each location using a Helios Hydra UX dual-beam plasma focused ion beam (PFIB)/SEM at an accelerating voltage of 20 keV, working distance of 4.0 mm, 0° tilt, and ×5000 magnification. The BSE images were cropped to 1024 × 1024 px for prediction by our segmentation models. EBSD data were acquired using a PFIB, equipped with an Oxford Synergy CMOS EBSD detector. EBSD data collection was carried out at an acceleration voltage of 20 keV, with a step size of 100 nm consistently applied across all measurements using AZtec software. AztecCrystal was used to generate inverse pole figure (IPF) maps, image quality (IQ) maps, and evaluate grain sizes. The mean equivalent circle diameter (ECD) was calculated for each image using a threshold angle of 5° and a minimum grain diameter of 0.4 μm.

Segmentation model

We used the UNet++ architecture[41] pretrained on either the ImageNet dataset, a general dataset commonly used for computer vision models, or on MicroNet, a large dataset of over 100,000 labeled microscopy images[26], to train semantic segmentation models to identify grain boundaries in SEM images of the FSP 316L stainless steel samples. Identifying grain boundaries in an SEM image is a highly class-imbalanced problem since grain boundaries class accounts for <10% of all pixels in an SEM image, where the rest is made of the grain class. A topological loss function, TopoLoss[42], was added to the DICE[43] loss function to train the models. We previously found that the incorporation of topological information during training improves the connectivity of the predicted grain boundary network[27]. The dual loss function is shown in Eq. 1.

$$\mathrm{Loss}_{\mathrm{TopoDICE}} = \mathrm{Loss}_{\mathrm{DICE}} + \lambda \mathrm{Loss}_{\mathrm{Topo}} \tag{1}$$

where $\mathrm{Loss}_{\mathrm{DICE}}$ is the DICE loss, $\mathrm{Loss}_{\mathrm{Topo}}$ is the topological loss, and λ is a hyperparameter that controls the relative weighting between the two losses. An ADAM optimizer with an adaptive learning rate from $1e^{-4}$ to $1e^{-5}$ was used to train the dual loss function.

Training set

Each registered SEM-EBSD pair (1792 × 1280 px) was split into three smaller rectangular patches of 256 × 1280 px to create three distinct test sets: test set 1 contains chips A1 to A5, test set 2 contains chips D1 to D5, and test set 3 contains chips F1 to F5. For each test set, the remainder of the image was split into non-overlapping square patches of 256 × 256 px to create the training set, as shown in Fig. 6. For predicting grain boundary maps of test set 1, chips in columns B to G were used as training data, while those in column A were used as test data. For predicting grain boundary maps of test set 2, chips in columns A to C and E to G were used as training data, while chips in column D were used as test data. For grain boundary maps of test set 3, image chips in columns A to E and G were used as training data, while chips in column F were used as test data. Details of the training and testing split are provided in supplementary information Table S1. Training three models with different dataset splits allowed us to (1) perform 3-fold cross-validation for each image modality and (2) generate an ensemble of models trained on each modality. During training, standard augmentation methods were employed by applying horizontal flips, vertical flips, and 90° rotations to the patches. Each augmented training set was then divided into 90%/10% train/validation splits. A total of 105 training and 11 validation sample patches, along with the aforementioned rectangular test sample, were obtained for each set.

Validation metrics

The F_1 score is a commonly used metric to measure the performance of supervised semantic segmentation models. The F_1 score is the harmonic mean of precision and recall and is computed as follows:

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{2}$$

where *TP* is the number of true positives (correctly classified pixels), *FP* is the number of false positives (pixels incorrectly classified as belonging to the grain boundary), and *FN* is the number of false negatives (pixels incorrectly classified as not belonging to the grain boundary). When we refer to a pixel belonging to a grain boundary, this indicates that the area contained within that pixel is adjacent to and/or overlapping with the physical grain boundary. The thickness (pixel width) of what we call the grain boundary is, therefore, arbitrary. False positives in pixel-wise segmentation due to thicker boundaries are not as meaningful as false positives away from grain boundary regions.

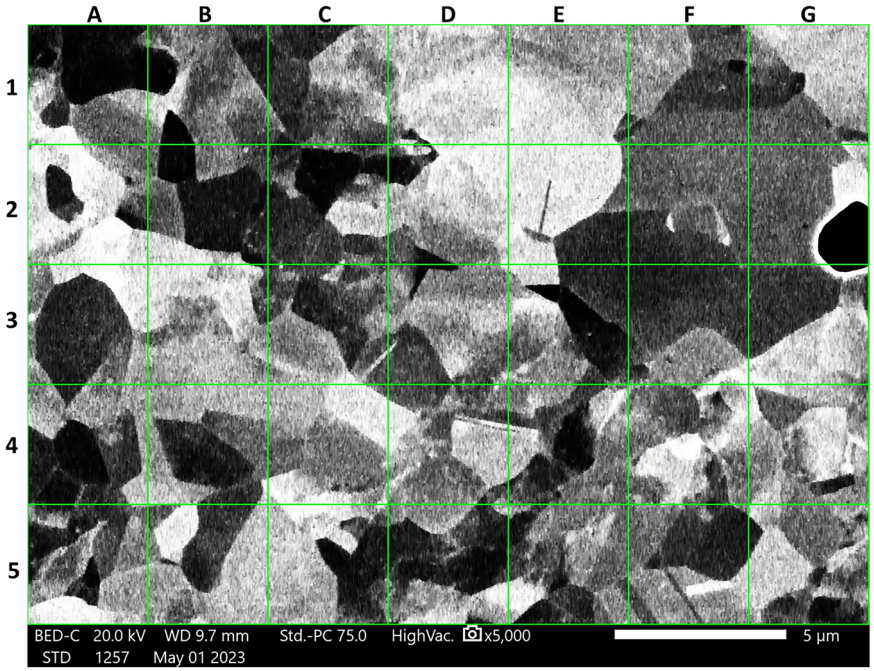

Fig. 6 | Location of the training set chips. The green lines demarcate individual chips. Columns A, D, and F were used as test sets in splits 1, 2, and 3, respectively. The remainder of the chips not in the corresponding test set were used for training.

In addition to the F_1 score as a general measure of model performance, it is useful to quantify the extent of the segmentation error. The Hausdorff distance provides such a measure by capturing the maximum discrepancy between two corresponding images (i.e., the segmentation map and the grain boundary map). In our application, the Hausdorff distance represents the maximum distance of an *FP* prediction to the nearest grain boundary pixel. Because the Hausdorff distance is highly sensitive to outliers, we applied the 95% percentile Hausdorff distance (HD95), which excludes the top 5% of distances that may largely contain these outliers[44].

The scikit-image[45] library was used to extract grain size and GB information from the grain boundary maps. Values expressed in pixels were converted to micrometers using a micron-to-pixel conversion factor (C) specific to the measurement. For the grain boundary map produced by EBSD, C was derived from the EBSD measurement. For the grain boundary maps aligned with BSE and SE micrograph, C was derived from the micrograph, using the micron-to-pixel conversion factor from the micrograph. When a grain boundary map was not continuous, which is often the case with predicted segmentation maps[22], automated grain detection and grain area measurements were inaccurate.

Data availability

Data used in this study are available to the corresponding author upon reasonable request.

Code availability

The codebase used for training semantic segmentation models is available on GitHub at https://github.com/nasa/pretrained-microscopy-models.

Received: 17 December 2024; Accepted: 8 September 2025;
Published online: 29 October 2025

References

1. Mishra, R. S. & Ma, Z. Y. Friction stir welding and processing. *Mater. Sci. Eng. R Rep.* **50**, 1–78 (2005).
2. Mishra, R. S., De, P. S. & Kumar, N. *Friction Stir Welding and Processing: Science and Engineering*, 13–58 (Springer, 2014).
3. Heidarzadeh, A. et al. Friction stir welding/processing of metals and alloys: a comprehensive review on microstructural evolution. *Prog. Mater. Sci.* **117**, 100752 (2021).
4. Mironov, S., Sato, Y. S., Kokawa, H., Inoue, H. & Tsuge, S. Structural response of superaustenitic stainless steel to friction stir welding. *Acta Mater.***59**, 5472–5481 (2011).
5. Liu, F. C. & Nelson, T. W. In-situ material flow pattern around probe during friction stir welding of austenitic stainless steel. *Mater. Des.* **110**, 354–364 (2016).
6. Wang, D. et al. Microstructural evolution and mechanical properties of friction stir welded joint of Fe–Cr–Mn–Mo–N austenite stainless steel. *Mater. Des.* **64**, 355–359 (2014).
7. Jiang, X., Overman, N., Canfield, N. & Ross, K. Friction stir processing of dual certified 304/304L austenitic stainless steel for improved cavitation erosion resistance. *Appl. Surf. Sci.* **471**, 387–393 (2019).
8. Chen, Y. C. et al. Friction stir processing of 316L stainless steel plate. *Sci. Technol. Weld. Join.* **14**, 197–201 (2009).
9. Hajian, M. et al. Microstructure and mechanical properties of friction stir processed AISI 316L stainless steel. *Mater. Des.* **67**, 82–94 (2015).
10. Sutton, B. et al. *Energy Materials 2017* (eds Xingbo, L. et al.) 343–351 (Springer, 2017).
11. Hansen, N. Hall–Petch relation and boundary strengthening. *Scr. Mater.***51**, 801–806 (2004).
12. Garcia, D. et al. In-situ measurement and control of the tool-workpiece interface temperature during friction stir processing of 304/304L stainless steel. *Mater. Today Commun.* **38**, 107672 (2024).
13. Wang, T., Garcia, D., Pole, M. & Ross, K. A. Force reduction of friction stir welding and processing of steel. *Materialia* **33**, 102050 (2024).
14. Liu, F. C. & Nelson, T. W. In-situ grain structure and texture evolution during friction stir welding of austenite stainless steel. *Mater. Des.* **115**, 467–478 (2017).
15. Mironov, S., Sato, Y. S. & Kokawa, H. Microstructural evolution during friction stir welding of Ti–15V–3Cr–3Al–3Sn alloy. *Mater. Sci. Eng. A* **527**, 7498–7504 (2010).
16. Mahmoudiniya, M. & Kestens, L. A. I. Microstructural development and texture evolution in the stir zone and thermomechanically affected zone of a ferrite-martensite steel friction stir weld. *Mater. Charact.* **175**, 111053 (2021).
17. Jeon, J. et al. Friction stir spot welding of single-crystal austenitic stainless steel. *Acta Mater.***59**, 7439–7449 (2011).
18. Ede, J. M. Deep learning in electron microscopy. *Mach. Learn. Sci. Technol.* **2**, 011004 (2021).

19. Botifoll, M., Pinto-Huguet, I. & Arbiol, J. Machine learning in electron microscopy for advanced nanocharacterization: current developments, available tools and future outlook. *Nanoscale Horiz.* **7**, 1427–1477 (2022).
20. Han, Y., Griffiths, R. J., Yu, H. Z. & Zhu, Y. Quantitative microstructure analysis for solid-state metal additive manufacturing via deep learning. *J. Mater. Res.* **35**, 1936–1948 (2020).
21. Morgan, D. et al. Machine learning in nuclear materials research. *Curr. Opin. Solid State Mater. Sci.* **26**, 100975 (2022).
22. Roberts, G. et al. Deep learning for semantic segmentation of defects in advanced STEM images of steels. *Sci. Rep.* **9**, 12744 (2019).
23. Patrick, M. J. et al. Automated grain boundary detection for bright-field transmission electron microscopy images via U-Net. *Microsc. Microanal.* **29**, 1968–1979 (2023).
24. Shen, C. et al. A generic high-throughput microstructure classification and quantification method for regular SEM images of complex steel microstructures combining EBSD labeling and deep learning. *J. Mater. Sci. Technol.* **93**, 191–204 (2021).
25. Hirabayashi, Y. et al. Deep learning for three-dimensional segmentation of electron microscopy images of complex ceramic materials. *NPJ Comput. Mater.* **10**, 46 (2024).
26. Stuckner, J., Harder, B. & Smith, T. M. Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset. *NPJ Comput. Mater.* **8**, 200 (2022).
27. Chowdhury, S. A. et al. Automated grain boundary (GB) segmentation and microstructural analysis in 347H stainless steel using deep learning and multimodal microscopy. *Integr. Mater. Manuf. Innov.* **13**, 244–256 (2024).
28. Goldstein, J. I. et al. *Scanning Electron Microscopy and X-ray Microanalysis* (Springer, 2017).
29. Lloyd, G. E. Atomic number and crystallographic contrast images with the SEM: a review of backscattered electron techniques. *Mineral. Mag.* **51**, 3–19 (1987).
30. Chen, D., Chang, C. P. & Loretto, M. H. Orientation contrast of secondary electron images from electropolished metals. *Ultramicroscopy* **156**, 41–49 (2015).
31. Joy, D. C. Beam interactions, contrast and resolution in the SEM. *J. Microsc.* **136**, 241–258 (1984).
32. Baskaran, A., Kane, G., Biggs, K., Hull, R. & Lewis, D. Adaptive characterization of microstructure dataset using a two stage machine learning approach. *Comput. Mater. Sci.* **177**, 109593 (2020).
33. Winiarski, B. et al. Correction of artefacts associated with large area EBSD. *Ultramicroscopy* **226**, 113315 (2021).
34. Tong, V. S. & Ben Britton, T. TrueEBSD: correcting spatial distortions in electron backscatter diffraction maps. *Ultramicroscopy* **221**, 113130 (2021).
35. Nowell, M. M., Witt, R. A. & True, B. W. EBSD sample preparation: techniques, tips, and tricks. *Microsc. Today* **13**, 44–49 (2018).
36. Sytwu, K., Rangel DaCosta, L. & Scott, M. C. Generalization across experimental parameters in neural network analysis of high-resolution transmission electron microscopy datasets. *Microsc. Microanal.***30**, 85–95 (2024).
37. Humphreys, F. J. Review Grain and subgrain characterisation by electron backscatter diffraction. *J. Mater. Sci.* **36**, 3833–3854 (2001).
38. Bachmann, F., Hielscher, R. & Schaeben, H. Grain detection from 2d and 3d EBSD data—specification of the MTEX algorithm. *Ultramicroscopy* **111**, 1720–1733 (2011).
39. Jazaeri, H. & Humphreys, F. J. Quantifying recrystallization by electron backscatter diffraction. *J. Microsc.* **213**, 241–246 (2004).
40. Patala, S., Mason, J. K. & Schuh, C. A. Improved representations of misorientation information for grain boundary science and engineering. *Prog. Mater. Sci.* **57**, 1383–1425 (2012).
41. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N. & Liang, J. Unet++: A nested u-net architecture for medical image segmentation. *Lect. Notes Comput. Sci* **11045**, 3–11 (2018).
42. Hu, X., Li, F., Samaras, D. & Chen, C. Topology-preserving deep image segmentation. *Adv. Neural Inf. Processing Syst.* **32**, https://proceedings.neurips.cc/paper_files/paper/2019/file/2d95666e2649fcfc6e3af75e09f5adb9-Paper.pdf (2019).
43. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. & Jorge Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Deep Learn Med. Image Anal. Multimodal Learn Clin. Decis. Support***2017**, 240–248 (2017).
44. Jia, J., Staring, M. & Stoel, B. C. seg-metrics: a Python package to compute segmentation metrics. Preprint at *medRxiv* https://doi.org/10.1101/2024.02.22.24303215 (2024).
45. Van der Walt, S. et al. scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

Acknowledgements

The research described in this paper is part of the Materials Characterization, Prediction, and Control agile investment at Pacific Northwest National Laboratory. It was conducted under the Laboratory Directed Research and Development Program at PNNL, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RL01830.

Author contributions

M.F.N.T. and J.N. trained and evaluated segmentation models. J.E. collected and analyzed the paired EBSD and SEM images. M.P. collected out-of-domain SEM images. K.N. and D.B. supported data organization and distribution. D.G., T.W., H.D. and K.R. provided sample preparation. E.B. and E.S. were in charge of overall direction and planning. K.S.K. and D.R.T. helped guide project direction and provided material science domain knowledge. J.A.B. generated grain boundary maps from the EBSD and SEM pairs, performed data analysis of model performance, and supervised the project. M.F.N.T., J.N., J.E., and J.A.B. wrote the main manuscript text. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-025-01801-4.

Correspondence and requests for materials should be addressed to Keerti S. Kappagantula or Jenna A. Bilbrey.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© Battelle Memorial Institute 2025